

Stat 225 - Review for exam 2

The second exam will be this coming Wednesday, October 24. Here are some problems that might help. Note that you may use a simple calculator.

1. When I execute the following Python code:

```
import pandas as pd
from numpy.random import seed
seed(75)
df = pd.read_csv("https://www.marksmath.org/data/county_small.csv")
df.sample(5)
```

I generate the following output:

state	name	FIPS	pop2010	hs_grad	bachelors
Kentucky	Hart County	21099	18199	67.7	9.2
North Carolina	Buncombe County	37021	238318	87.2	31.2
Missouri	Polk County	29167	31137	79.5	16.6
Texas	Lamar County	48277	49793	82.4	17.4
Indiana	Whitley County	18183	33292	90.2	16.6

- (a) What are the cases?
 - (b) What are the variables?
 - (c) Classify each variable as numerical or categorical.
 - (d) Further classify each numerical variable as continuous or discrete.
 - (e) Further classify each categorical variable as nominal or ordinal.
 - (f) What is the purpose of the sample command (i.e. `df.sample(5)`) at the end?
2. A sample of 100 North Carolina adults finds their average height to be 67.88 inches with a standard deviation of 4.12 inches
 - (a) What is the standard error associated with this sample mean?
 - (b) Write down a 95% confidence interval for the height of a North Carolina adult based on this sample.
 3. A sample of 10 North Carolina adults finds their average height to be 66.7 inches with a standard deviation of 4.8 inches
 - (a) What is the standard error associated with this sample mean?
 - (b) Write down a 95% confidence interval for the height of a North Carolina adult based on this sample.

4. A sample of 100 North Carolina adults finds that 77 of them exercise some.
 - (a) What is the standard error associated with this sample proportion?
 - (b) Write down a 95% confidence interval for the height of a North Carolina adult based on this sample.
5. A sample of 10 North Carolina adults finds that 9 of them exercise some.
 - (a) What is the standard error associated with this sample proportion?
 - (b) Write down a 95% confidence interval for the height of a North Carolina adult based on this sample.
6. Suppose we would like to use the data from problem 2 to explore the following question:
Is the average North Carolinian 69.5 inches tall.
 - (a) Write down the hypothesis test that clearly states the problem.
 - (b) Compute the p -value obtained from this data.
 - (c) What is the conclusion of the hypothesis test?
7. Repeat the previous problem using the data from problem 3.
8. Suppose we would like to use the data from problem 4 to explore the following question:
Do $3/4$ of North Carolinians exercise some.
 - (a) Write down the hypothesis test that clearly states the problem.
 - (b) Compute the p -value obtained from this data.
 - (c) What is the conclusion of the hypothesis test?
9. Repeat the previous problem using the data from problem 5.
10. I'd like to design a study to determine the proportion of North Carolinians who smoke and I'd like to obtain a 99% confidence interval whose width is 0.02. How many people should I include in my study?